Le DiCo et sa version DiCouèbe

Document descriptif et manuel d'utilisation

Version du rapport : 1.0 — 19 avril 2005



Document préparé par Anne-Laure Jousse et Alain Polguère
Observatoire de linguistique Sens-Texte (OLST)
Université de Montréal

Table des matières

1 Préambule

1.1 Ressources présentées dans ce document

Le **DiCo** (officiellement, DiCo-OLST) est une base de données lexicale du français, développée depuis plusieurs années à l'Observatoire de linguistique Sens-Texte (OLST) de l'Université de Montréal par Igor Mel'čuk et Alain Polguère. Le but du projet DiCo est la construction d'un modèle formel du lexique noyau du français, modèle à partir duquel peuvent être dérivés deux types de produits lexicographiques. Premièrement, on doit pouvoir produire automatiquement à partir du DiCo (par compilation de celui-ci) des lexiques de systèmes de traitement automatique de la langue (TAL). Deuxièmement, le DiCo doit contenir toute l'information nécessaire pour en dériver des versions « vulgarisées » à l'usage du grand public, dans un dictionnaire comme le *Lexique actif du français* (LAF) — voir Polguère (2000).

Le **DiCouèbe** est une interface permettant de consulter en ligne une version compilée du DiCo. Cette consultation s'effectue à l'aide d'un simple navigateur, directement à partir du site de l'OLST, à l'adresse suivante : http://www.olst.umontreal.ca/dicouebe.

1.2 Remerciements

1.2.1 Projet DiCo

Les personnes suivantes ont contribué, en tant qu'étudiants auxiliaires de recherche ou stagiaires, au projet DiCo : Dominic Anctil, Joel Altman, Sabrina Ben Hamou, Aline Francœur, Catherine Gagnon, Andréanne Germain-Rouleau, Bénédicte Grizolle, Anne-Laure Jousse, François Lareau, Sara-Anne Leblanc, Suzanne Mantha, Jasmina Milićević, Sylvain Neuvel, Mylène Perrault, Stéfan Popovic et Ophélie Tremblay.

Notre collègue Sylvain Kahane (Lattice/Modyco, Université Paris 10) nous a fourni de précieux conseils pour améliorer notre travail. Lidija Iordanskaja (OLST — Université de Montréal) a activement participé à l'élaboration d'un grand nombre de fiches DiCo.

1.2.2 Projet DiCompil-DiCoSQL-DiCouèbe

Le compilateur de DiCo (DiCompil) et l'interface DiCouèbe ont été réalisés au laboratoire Lattice du CNRS dans le cadre d'une Chaire Blaise Pascal attribuée à Igor Mel'čuk et gérée par la Fondation de l'École Normale Supérieure sous la direction de Sylvain Kahane.

- Maîtrise d'ouvrage : Sylvain Kahane, Alain Polguère, Jacques Steinlin.
- Chef de projet : Jacques Steinlin (Lattice, Université Paris 7).
- Collaborateurs : Aurélien Amacker, Florence Berbain, Adil El Ghali, Cédric Lecour, Philippe Lorin, Stéfan Popovic, Fabien Tarissan.

Remerciements également à Sébastien Cabot, Patrick Drouin, Anne-Laure Jousse, Alexis Nasr, Laurence Perrin et bien sûr Igor Mel'čuk.

Avertissement. Ce document n'est ni un texte scientifique ni une présentation détaillée du DiCo et de sa version en ligne DiCouèbe. Il s'agit d'une courte documentation, qui vise simplement à donner un aperçu de ce que l'utilisateur pourra trouver comme informations sur le lexique français grâce au DiCouèbe. On pourra se reporter aux textes de la bibliographie pour un approfondissement de l'approche lexicographique sous-jacente au DiCo ainsi que pour des compléments d'information sur le projet lui-même. Plusieurs textes sont par ailleurs disponibles sur le site de l'OLST :

http://www.olst.umontreal.ca [voir le menu Ressources>Textes à télécharger]

Le présent document sera régulièrement mis à jour.

2 Présentation générale du DiCo

Le DiCo a pour arrière-plan théorique la **lexicologie explicative et combinatoire** (**LEC**), la composante lexicologique de la théorie Sens-Texte, telle que présentée dans Mel'čuk et autres (1995). Le DiCo est en quelque sorte une version simplifiée et informatisée du *Dictionnaire explicatif et combinatoire du français contemporain* (DEC), publié dans Mel'čuk et autres (1984, 1988, 1992, 1999). Il se démarque du DEC dans la mesure où il est centré sur la description des **liens lexicaux paradigmatiques et syntagmatiques** contrôlés par les **lexies** (= unités lexicales) de la langue, plutôt que sur la description de leur sens. La base ne contient donc pas de véritable définition lexicale ; on y trouve cependant un étiquetage sémantique des lexies (voir section 3.3) ainsi que la description de leur structure actancielle.

2.1 Deux types de liens lexicaux

Cette section décrit la nature des deux types de liens lexicaux encodés dans le DiCo : liens paradigmatiques *vs* liens syntagmatiques.

> La section 2.1.1 ci-dessous est empruntée à l'introduction du *Lexique actif du français* — Mel'čuk et Polguère (en préparation) ; la section 2.1.2 contient du matériau extrait de plusieurs publications, comme Polguère (2003a) et Polguère (à paraître).

2.1.1 Liens paradigmatiques

Selon l'approche de la LEC, un lien paradigmatique connecte la lexie L2 à la lexie L1 (L1→L2) si L2 est un dérivé sémantique de L1. La notion traditionnelle de dérivation est bien connue. Elle correspond à des relations à la fois sémantiques et formelles entre deux mots, relations du type *fumer* → *fumeur* (nom d'agent), *compter* → *comptage* (nom d'action), *noir* → *noircir* (verbe causatif), etc. La LEC, cependant, considère aussi une notion de dérivation plus large, à savoir la **dérivation sémantique**.

Une dérivation sémantique est une relation entre deux lexies basée sur une parenté de sens. Plus précisément, une lexie L2 est dite sémantiquement dérivée d'une lexie L1 si et seulement si les trois conditions suivantes sont remplies :

1. L2 entretient une relation sémantique avec L1, c'est-à-dire que les définitions de L1 et L2 possèdent des composantes de sens communes. Dans le cas le plus typique, L2 se définit en terme de L1. Par exemple, la lexie HACHE [= L2] est définie en terme de COUPER [= L1], car *hache* signifie 'artefact servant à **couper**...'.
2. La relation sémantique entre L2 et L1 est récurrente dans la langue. Par exemple, la relation entre HACHE et COUPER, 'artefact servant à...', est récurrente en français : FRAPPER → MARTEAU, OUVRIR → CLÉ, FUMER → PIPE, etc.

3. La relation entre L1 et L2 s'exprime éventuellement, mais pas nécessairement, de façon morphologique dans la langue. Par exemple, pour la relation 'artefact servant à…', on trouve en français les dérivations suffixales suivantes : *bouch*(*-er*) → *bouch+on*, *ras*(*-er*) → *ras+oir*, *décapsul*(*-er*) → *décapsul+eur*, etc.

Puisque la relation sémantique qui lie HACHE à COUPER satisfait les trois conditions ci-dessus, nous pouvons dire que la lexie HACHE est sémantiquement dérivée de COUPER.

> Un lien de dérivation sémantique qui est marqué explicitement par un moyen morphologique correspond à une dérivation au sens traditionnel. La dérivation au sens traditionnel est donc un cas particulier de dérivation sémantique.

On peut isoler trois grandes familles de dérivations sémantiques, qui correspondent aux trois cas de figure ci-dessous.

1. Les deux lexies possèdent (approximativement) le même sens. Il s'agit ici d'une dérivation sémantique (quasi) vide, qui correspond aux cas bien connus suivants : (i) synonymie exacte ou approximative (*vélo* est sémantiquement équivalent à *bicyclette* ; *défendre* est sémantiquement presque équivalent à *interdire*) ; (ii) conversion exacte ou approximative (*X précède Y* est équivalent à *Y suit X* ; *X donne Y à Z* est presque équivalent à *Z reçoit Y de X*) ; changement de partie du discours (*marchandage* est le nom d'action pour *marchander*, *vente* pour *vendre*, *coup* pour *frapper* ; *terrestre* est l'adjectif pour [*relatif à la*] *terre*, *paternel* pour [*relatif au*] *père*, *urbain* pour [*relatif à la*] *ville*). Par convention, on inclut également dans cette première famille les termes génériques (*cours d'eau* pour *rivière*, *légume* pour *asperge*…).
2. Les deux lexies possèdent des sens opposés (*interdire* vs *autoriser*, *petit* vs *grand*…). Il s'agit ici de l'antonymie exacte ou approximative.
3. Une des deux lexies désigne un élément de la situation désignée par l'autre. Il peut s'agir d'un participant (*nager* → *nageur*, *envoyer* [*qqch. par la poste*] → *destinataire*, *accident* → *victime*…), d'un circonstant (*dormir* → *lit*, *patiner* → *patinoire*…) ou d'une caractéristique d'un participant ou circonstant (*s'irriter* → *irritable*, *comprendre* → *compréhensible*…).

2.1.2 Liens syntagmatiques

Un lien syntagmatique connecte la lexie L2 à la lexie L1 si l'expression L1+L2 est une **collocation** dont L1 est la base et L2 le collocatif. Voici une brève présentation des notions de collocation, base et collocatif.

Une collocation est une expression *L1L2* (ou *L2L1*), formée des lexies L1 et L2, que le locuteur produit d'une façon particulière : il sélectionne L1 librement d'après son sens 'L1', alors qu'il sélectionne L2 pour exprimer un sens 's' en fonction de L1. La collocation est donc une expression semi-idiomatique. Voici trois exemples de collocations du français :

- *grosse*[=L2] *tempête*[=L1]
- *dormir* [=L1] *profondément* [=L2]
- *pleuvoir* [=L1] *des cordes* [=L2]

On appelle **base** de la collocation l'élément qui, sélectionné librement par le locuteur, retient son sens dans la collocation et la contrôle. Dans les exemples ci-dessus, les bases des collocations sont les éléments étiquetés L1. Les éléments étiquetés L2 sont appelés **collocatifs** de la base (dans la collocation).

2.2 ENCODAGE DES LIENS PAR LES FONCTIONS LEXICALES

2.2.1 Fonctions lexicales standard

Les liens lexicaux sont encodés dans le DiCo au moyen des fonctions lexicales standard de la théorie Sens-Texte. Voici une brève présentation de la notion de fonction lexicale, en partie empruntée à Polguère (à paraître). Pour une présentation détaillée de la notion de fonction lexicale, voir Mel'čuk et autres (1995:125-152) et Mel'čuk (1996).

Une **fonction lexicale standard** — par exemple, **Syn**, **Anti**, **V**$_0$, **S**$_0$, **Magn**, **Oper**$_i$, etc. — est une entité lexicale à caractère universel (c'est-à-dire, présente dans toutes les langues naturelles) qui se caractérise par deux paramètres : (i) un sens très vague et (ii) une configuration syntaxique d'emploi. Chaque fonction lexicale **f** doit être appliquée à une lexie donnée L pour retourner un ensemble de lexies dont chacune est associée au sens et à la configuration syntaxique d'emploi de **f** en regard de L. La fonction lexicale est donc une généralisation de la notion de lexie ; c'est en quelque sorte un « potentiel de lexies ».

Au niveau de la formalisation, il faut mentionner les deux points suivants, illustrés à partir de la fonction lexicale de synonymie **Syn**.

1. L'application de **f** à L est notée **f**(*L*) — par exemple, l'expression « **Syn**(*réfrigérateur*) » désigne les synonymes exacts de la lexie RÉFRIGÉRATEUR. L est appelé l'**argument** de **f** dans l'application **f**(*L*) ; l'ensemble des lexies retourné par **f**(*L*) est appelé la **valeur** de **f**(*L*).
2. On décrit les liens lexicaux de type **f** qui vont de L à chacun des éléments de la valeur v de **f**(*L*) en posant l'égalité « **f**(*L*) = v ». Par exemple, la formule :

   Syn(*réfrigérateur*) = *frigidaire*, fam. *frigo*

   décrit l'ensemble des synonymes (exacts) de RÉFRIGÉRATEUR.

Le système des fonctions lexicales standard étant utilisé pour modéliser et encoder les liens lexicaux récurrents de la langue, on distingue deux familles de fonctions lexicales — qui sont parallèles aux deux types de liens lexicaux présentés plus hauts.

1. Les **fonctions lexicales paradigmatiques**, comme **`Syn`**, retournent des valeurs qui entretiennent des liens paradigmatiques (c'est-à-dire coupant à travers le lexique) avec l'argument de la fonction.
2. Les **fonctions lexicales syntagmatiques** retournent quant à elles des valeurs qui entretiennent des liens de combinatoire avec l'argument de la fonction : ce sont des collocatifs au sein de collocations dont l'argument de la fonction est la base.

Il existe une soixantaine de fonctions lexicales standard, en voici quelques exemples[1].

<u>Exemples de fonctions lexicales paradigmatiques</u>

- **Conv** [conversif] — par ex., $\mathbf{Conv}_{321}$(*vendre*) = *acheter*
- **Anti** [antonyme] — par ex., **Anti**(*bien*) = *mal*
- **Contr** [contrastif] — par ex., **Contr**(*mer*) = *terre*
- **Sing** [nom d'unité régulière] — par ex., **Sing**(*neige*) = *flocon*

<u>Exemples de fonctions lexicales syntagmatiques</u>

- **Magn** [intensificateur] — par ex., **Magn**(*féliciter*) = *chaleureusement*, *vivement*
- $\mathbf{Oper}_i$, $\mathbf{Func}_i$, $\mathbf{Labor}_{ij}$ [types particuliers de verbes supports] — par ex. :

$\mathbf{Oper}_1$(*haine*)	=	*avoir*, *concevoir*, *éprouver*, *nourrir*, *ressentir* [ART ~ Prép_{envers} N_Y], *porter*, *vouer* [ART ~ *à* N_Y]; *entretenir* [ART ~ Prép_{envers} N_Y] | avec modificateur [*Il entretient une haine tenace envers son père.*]
$\mathbf{Func}_1$(*haine*)	=	*animer*, *habiter*, *remplir* [N_X]
$\mathbf{Labor}_{12}$(*haine*)	=	*avoir* [N_Y *en* ~]

2.2.2 Fonctions lexicales non standard

Outre les fonctions lexicales standard, on utilise dans le DiCo (et dans la LEC en général) des fonctions lexicales non standard pour représenter des relations qui, notamment, ne sont pas considérées comme possédant une validité linguistique « universelle » ou ne sont pas récurrentes dans la langue.

Par exemple, on trouve dans le DiCo, pour la première acception nominale de CRIMINEL, les liens de fonctions lexicales non standard suivants :

[1] Une liste des fonctions lexicales standard est donnée en annexe.

CRIMINEL	I	Qui cherche à échapper à la justice	_en cavale_
CRIMINEL	I	Qui cherche à échapper à la justice	_en fuite_
CRIMINEL	I	C. reconnaît avoir commis Y	avouer
CRIMINEL	I	C. reconnaît avoir commis Y	se dénoncer
CRIMINEL	I	C. éprouve du regret pour Y	se repentir
CRIMINEL	I	C. commet un autre crime	récidiver
CRIMINEL	I	Celui qui C. commet un autre crime	récidiviste

Cette présentation est bien entendu très superficielle. On pourra consulter Polguère (sous presse) pour une étude détaillée des notions de fonction lexicale standard *vs* non standard.

2.3 Actants sémantiques

La notion d'actant sémantique est présentée de la façon suivante dans les pages d'introduction du LAF — Mel'čuk et Polguère (à paraître).

> Dans la plupart des cas, le sens d'une lexie ne peut être clairement compris, et donc décrit, sans prendre en compte les « participants » de la situation que cette lexie désigne. Nous appellerons ceux-ci **actants sémantiques** et nous les identifierons dans les articles du LAF par les variables *X*, *Y*...

Dans le DiCo, chaque lexie décrite est présentée avec sa **forme propositionnelle**, c'est-à-dire avec une formule explicitant le nombre d'actants qu'elle possède. Par exemple :

COLLER	I.1a	: L'individu X ~ L'objet Y À L'objet Z AVEC LA substance W

Il faut conclure ici cette très rapide revue du type de propriétés lexicales décrites dans le DiCo pour passer à la présentation de la base de données elle-même.

3 Structure du DiCo

3.1 Nomenclature

La nomenclature initiale du DiCo est très sélective. Elle se focalise avant tout sur la description des lexies ayant les trois caractéristiques suivantes :

1. ce sont des lexies courantes du français ;
2. elles contrôlent un certain nombre de dérivations sémantiques ou de collocations, ce qui en rend la description plus pertinente dans le cadre d'une base de type DiCo ;
3. elles forment ensemble une sorte de noyau lexical de la langue — un « français fondamental ».

La version du DiCo actuellement en ligne contient la description de quelques centaines de **vocables** (mots polysémiques), ce qui correspond à l'encodage d'environ 20,000 liens lexicaux. (Voir les statistiques à jour sur la page d'accueil du DiCouèbe.) Il s'agit d'une première étape. Une nomenclature cible de 3000 vocables a été constituée. Sa description correspondra à la seconde étape majeure de développement de la base.

Dans le processus de rédaction du DiCo, on distingue plusieurs niveaux d'avancement. Le **statut 3** signifie que la lexie qui est décrite dans la fiche en question n'a pas encore été travaillée. Les données linguistiques présentes dans la fiche sont très sommaires. Le **statut 2** signifie que la fiche contient des informations linguistiques, mais que leur traitement est toujours en cours. Le **statut 1** indique que la fiche est dégrossie, mais qu'il manque une validation des données linguistiques et de leur modélisation. Enfin, le **statut 0** signifie que le traitement est complet. Seul le contenu des fiches de statut 0 est accessible par le DiCouèbe.

3.2 Séparation des acceptions

Le DiCo est développé sous forme de base de données FileMaker Pro®. Chaque vocable est modélisé par un ensemble de fiches, chacune correspondant à une acception (une lexie) du vocable. Un vocable monosémique ne possède donc qu'une fiche. Un vocable polysémique, par contre, regroupe autant de fiches qu'il a d'acceptions, **plus** une fiche où sont stockées les propriétés partagées par toutes les lexies du vocable (généralement, la partie du discours, le genre du nom, etc.) ; cette fiche additionnelle porte le numéro d'acception `0`.

> Toutes les fiches des lexies d'un même vocable doivent avoir le même statut (0, 1, 2 ou 3).

La lexie décrite dans une fiche DiCo est appelée le **mot-clé** de la fiche. Voici, à titre d'exemple, l'énumération des fiches associées aux vocables homonymes FACTEUR(1), FACTEUR(2) et FACTEUR(3) dans le DiCo.

FACTEUR(1)		Le facteur a la clé de ma boîte aux lettres.
FACTEUR(2)		Outre les organistes, on a invité des facteurs d'orgue et des enseignants.
FACTEUR(3)	I	Le niveau de croissance sera un facteur déterminant pour le maintien de l'emploi.
FACTEUR(3)	II.1	Choisissez un écran solaire ayant un facteur de protection d'au moins 15.
FACTEUR(3)	II.2	Lorsqu'on veut comparer le rendement d'un certificat de placement garanti avec un rendement de dividende, il faut multiplier le rendement du dividende par le facteur 1,25.

Les chiffres entre parenthèses — ici (1), (2) et (3) — distinguent les vocables homonymiques. À l'intérieur d'un même vocable, la distance sémantique entre les lexies d'un même vocable est représentée par un système de numérotation. Les chiffres romains marquent une distance importante de sens entre lexies d'un même vocable (écarts métonymiques et métaphoriques non réguliers), les chiffres arabes une distance moindre et les lettres une très courte distance ou un type de lien de polysémie très régulier dans la langue. Par exemple, la numérotation apparaissant dans le tableau ci-dessus montre que, selon la modélisation du DiCo, la lexie FACTEUR(3)**II.1** est sémantiquement plus proche de FACTEUR(3)**II.2** qu'elle ne ne l'est de FACTEUR (3)**I.1**.

3.3 ÉTIQUETTES SÉMANTIQUES

Le DiCo ne propose pas de définition, toutefois, la caractérisation des sens lexicaux est partiellement prise en charge par un système d'**étiquettes sémantiques**. On trouvera dans Polguère (2003b) une description de la stratégie d'étiquetage sémantique utilisée dans le DiCo[2]. *Grosso modo*, une étiquette sémantique — `action`, `état`, `artefact`, etc. — permet de caractériser le type général de signification véhiculée par le mot-vedette. Cette étiquette est particulièrement utile pour distinguer les différentes acceptions regroupées sous un même vocable. Par exemple, le tableau ci-dessous donne les étiquettes sémantique de chacune des acceptions de BAGARRE, avec des phrases d'exemples pour illustrer les sens en question.

BAGARRE	a	échange de coups	Un gardien a été grièvement blessé en tentant de s'interposer dans une bagarre.
BAGARRE	b	confrontation verbale	L'innocente coutume de la galette des rois donne lieu à une bagarre épique entre les producteurs de fèves.

Formellement, une étiquette sémantique est un mot ou groupe de mots ayant une signification très générale et pouvant être utilisé comme le terme central dans la définition du mot-vedette (par exemple, BAGARRE**a** = 'échange de coups...'). L'ensemble des étiquettes sémantiques est organisé de façon hiérarchique, des étiquettes les plus générales (`fait` et `entité`) aux étiquettes les plus pointues. À remarquer que l'étiquette `individu` renvoie à un être humain alors que l'étiquette `personne` renvoie aussi bien à un être humain qu'à une entité sociale (syndicat, institution, etc.).

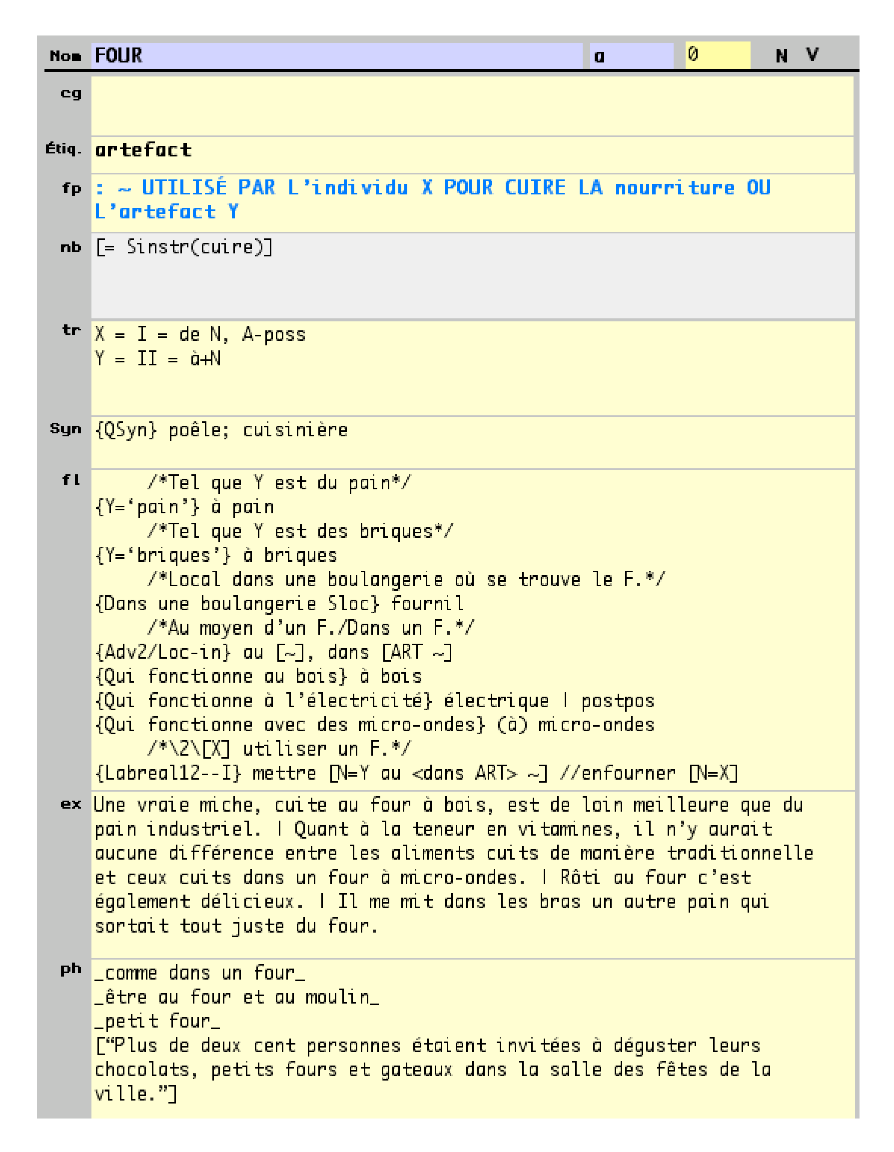
---

[2] Noter qu'il s'agit d'un texte qui n'est plus tout à fait à jour. Nous prévoyons de rendre prochainement disponible en ligne la hiérarchie d'étiquettes sémantiques du DiCo, accompagnée d'un document explicatif.

4 Le DiCo et sa version en ligne DiCouèbe

4.1 Le DiCo sous forme de fiches lexicographiques

Tel que mentionné plus haut, le DiCo est développé à l'heure actuelle forme d'une base de données grâce au logiciel FileMaker Pro®. Le format de base de données permet de séparer les champs d'une fiche et de les traiter de façon indépendante. Les différents champs du DiCo seront présentés plus bas, à la section 5. Voici, comme exemple, la fiche DiCo de la lexie FOUR**a** (dans la mise en page « Lexie », qui est celle visualisant le maximum de champs de nature linguistique).


Nom FOUR a 0 N V
cg
Étiq. artefact
fp : ~ UTILISÉ PAR L'individu X POUR CUIRE LA nourriture OU L'artefact Y
nb [= Sinstr(cuire)]
tr X = I = de N, A-poss
Y = II = à+N
Syn {QSyn} poêle; cuisinière
fl /*Tel que Y est du pain*/
{Y='pain'} à pain
/*Tel que Y est des briques*/
{Y='briques'} à briques
/*Local dans une boulangerie où se trouve le F.*/
{Dans une boulangerie Sloc} fournil
/*Au moyen d'un F./Dans un F.*/
{Adv2/Loc-in} au [~], dans [ART ~]
{Qui fonctionne au bois} à bois
{Qui fonctionne à l'électricité} électrique | postpos
{Qui fonctionne avec des micro-ondes} (à) micro-ondes
/*\2\[X] utiliser un F.*/
{Labreal12--I} mettre [N=Y au <dans ART> ~] //enfourner [N=X]
ex Une vraie miche, cuite au four à bois, est de loin meilleure que du pain industriel. | Quant à la teneur en vitamines, il n'y aurait aucune différence entre les aliments cuits de manière traditionnelle et ceux cuits dans un four à micro-ondes. | Rôti au four c'est également délicieux. | Il me mit dans les bras un autre pain qui sortait tout juste du four.
ph _comme dans un four_
être au four et au moulin
petit four
["Plus de deux cent personnes étaient invitées à déguster leurs chocolats, petits fours et gateaux dans la salle des fêtes de la ville."]

Ce format de données convient au travail lexicographique proprement dit, c'est-à-dire à l'**écriture** d'articles de DiCo. Nous allons maintenant voir comment se présente la version en ligne du DiCo, le DiCouèbe, dont la finalité actuelle est la **consultation** de la base.

4.2 Le DiCobjet et le DiCouèbe

Bien que le DiCo soit une base de données sous forme électronique, il possède une structure assez traditionnelle, inspirée des dictionnaires papier d'usage courant. Jacques Steinlin a développé, au laboratoire Lattice-Talana de Paris 7, une architecture beaucoup plus riche — le **DiCobjet** — permettant d'améliorer et affiner les différents modes d'accès aux données. Deux citations extraites de Steinlin et autres (2004) permettront de mieux faire comprendre l'écart séparant le DiCo de sa version en ligne.

> Cette architecture [= le DiCobjet] vise d'une part à améliorer les possibilités de consultation du DiCo par des utilisateurs humains et, par là même, à rendre les informations lexicographiques exploitables par des systèmes de traitement automatique de la langue (notamment en génération de textes). D'autre part, elle ambitionne de faciliter toutes les tâches lexicographiques proprement dites.
>
> Le contenu du DiCo est transmis au compilateur sous forme d'un texte tabulé où les articles sont à la suite les uns des autres. Le compilateur réalise une analyse syntaxique du fichier et produit une représentation sémantique, le DiCobjet. Le travail du compilateur consiste donc à identifier des atomes d'information et à les organiser dans une structure de données adéquate.

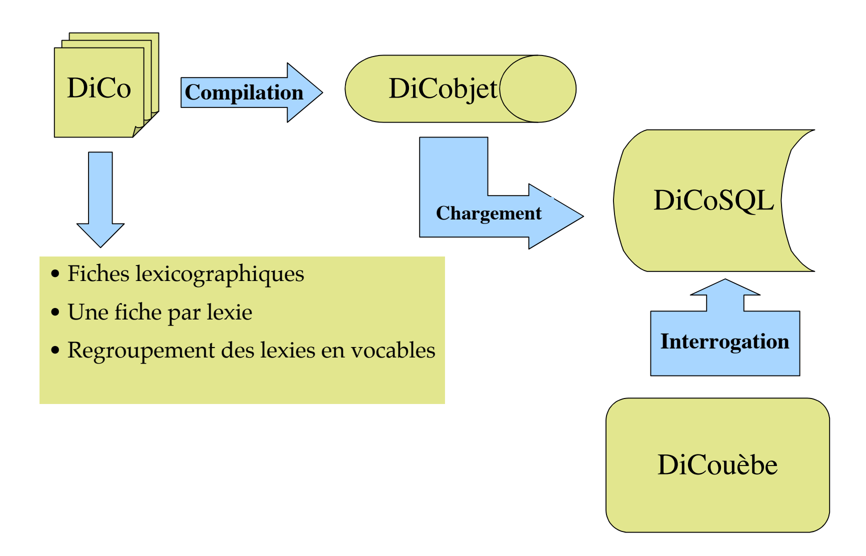


Figure 1 : DiCo, DiCobjet, DiCoSQL et DiCouèbe

Notons que nous utiliserons le nom **DiCouèbe** pour désigner non seulement l'interface d'accès au DiCo, mais aussi la base lexicale compilée, stockée dans le **DiCoSQL**.

Un mini-guide d'utilisation de l'interface du DiCouèbe est offert à la section 6 ; nous allons maintenant présenter les différents champs d'une fiche du DiCo et les informations qu'ils contiennent.

5 Les champs d'une fiche DiCo

Une fiche DiCo est divisée en 17 champs. Chaque champ est réservé à un type particulier d'information. Il est important de comprendre quelles données sont stockées dans le DiCo avant compilation, puisque ce sont ces données que l'on extrait, sous une forme « atomisée », au moyen de l'interface du DiCouèbe. Les illustrations présentées ci-dessous sont empruntées à la base DiCo, sous son format FileMaker Pro®.

5.1 Nom

Ce premier champ contient le nom du vocable dont relève la fiche.

5.2 Numéro d'acception de la lexie [•]

Ce champ indique l'acception du vocable décrite dans la fiche. Il est vide lorsque le vocable est monosémique. Rappel : une fiche contenant la valeur `0` dans ce champ ne correspond pas à une acception, mais contient les propriétés partagées par toutes les lexies du vocable.

5.3 Statut

Rappelons, voir section 3.1, que le statut indique le stade de traitement de la fiche. Il prend une valeur de `0` à `3`.

La sortie d'écran ci-dessous montre le contenu des trois champs qui viennent d'être présentés pour la fiche de FOUR**a**.



5.4 Remarques [Rem.]

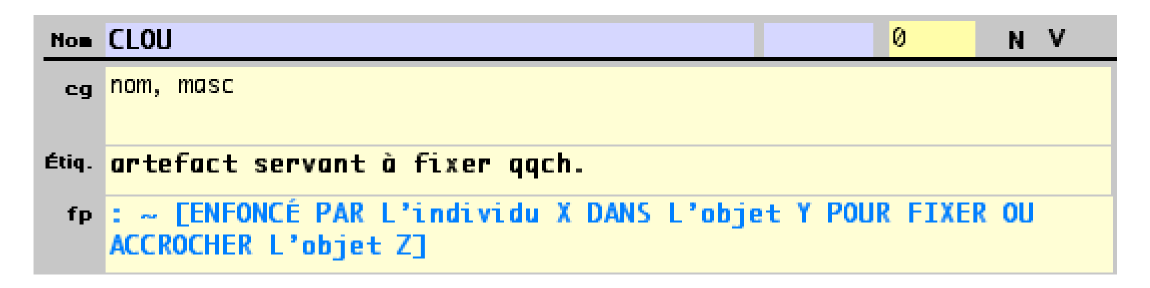
Les remarques stockées dans ce champ servent uniquement aux lexicographes du DiCo. Elles ne sont pas accessibles au moyen du DiCouèbe, pour les usagers qui ne se servent de l'interface qu'à fins de consultation.

5.5 Unité lexicale de base [ULB]

Ce champ peut prendre deux valeurs : `Oui` ou `Non`. Il indique si la lexie décrite est ou non l'**unité lexicale de base** d'un vocable polysémique. Pour une présentation de la notion d'unité lexicale de base et des principes permettant de dégager la structure polysémique d'un vocable, voir Mel'čuk et autres (1995). Noter que, dans le cas d'un vocable monosémique (c'est-à-dire ne contenant qu'une lexie), la valeur de ce champ est `Oui`.

5.6 CARACTÉRISTIQUES GRAMMATICALES [CG]

Ce champ comporte des informations d'ordre grammatical comme la partie du discours, le genre grammatical (pour les noms), les marques d'usage et autres restrictions d'emploi, etc. En voici un exemple :



5.7 ÉTIQUETTE SÉMANTIQUE [ÉTIQ.]

Les étiquettes sémantiques ont été présentées à la section 3.3. La sortie d'écran ci-dessus contient un exemple d'encodage, pour une étiquette assez atypique (`artefact servant à fixer qqch.`). Cette dernière est en effet formellement constituée d'une expression linguistique très complexe, presque équivalente à une définition.

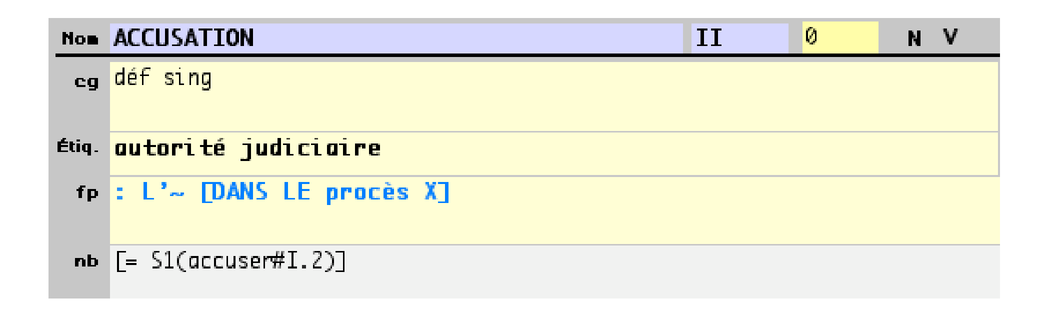
5.8 FORME PROPOSITIONNELLE [FP]

La notion de forme propositionnelle a déjà été introduite à la section 2.3. La sortie d'écran pour CLOU**a** ci-dessus donne un exemple d'encodage de ce type de formule dans le DiCo.

5.9 NOTA BENE [NB]

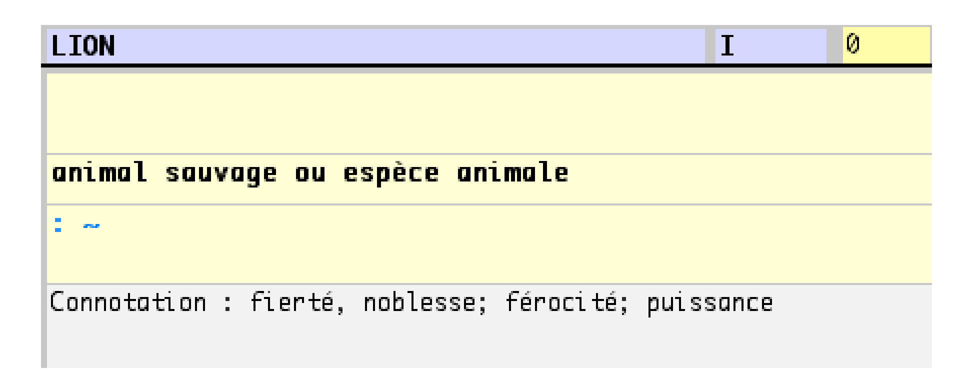
Ce champ sert à stocker certaines informations qui ont trait, dans la plupart des cas, aux liens lexicaux, mais ne peuvent être encodées directement par des liens de fonctions lexicales véritables. Voici trois types courants d'informations stockées dans le *Nota Bene.*

1. On peut trouver dans ce champ l'indication du fait que le mot-clé est une valeur d'une fonction lexicale d'une autre lexie. Il s'agit donc en quelque sorte d'un lien de fonction lexicale inverse ; par exemple :



Le séparateur « # » est utilisé dans le DiCo pour introduire les numéros d'acception des vocables. Ainsi, dans la sortie d'écran ci-dessus, `accuser#I.2` désigne la lexie numéro **I.2** du vocable ACCUSER.

2. Les **connotations** associées au mot-clé sont aussi introduites dans ce champ, plutôt que dans les champs dédiés aux liens lexicaux. En fait, la connotation peut être vue comme un lien entre une lexie et un « concept » associé à cette lexie, indépendamment des lexies ou expressions linguistiques permettant d'exprimer ce concept ; par exemple :

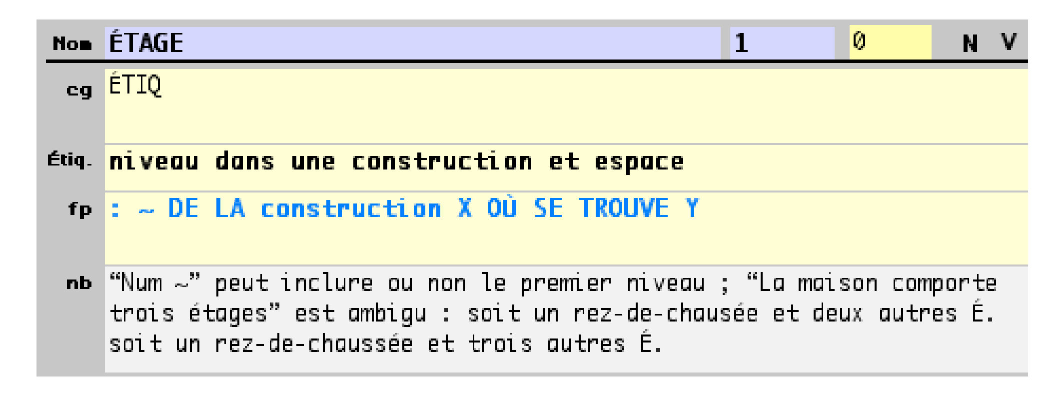

LION | I | 0

animal sauvage ou espèce animale

: ~

Connotation : fierté, noblesse; férocité; puissance

3. Finalement, ce champ peut contenir des remarques sur l'usage du mot-clé ; par exemple :


Nom ÉTAGE | 1 | 0 | N V

cg ÉTIQ

Étiq. niveau dans une construction et espace

fp : ~ DE LA construction X OÙ SE TROUVE Y

nb "Num ~" peut inclure ou non le premier niveau ; "La maison comporte trois étages" est ambigu : soit un rez-de-chaussée et deux autres É. soit un rez-de-chaussée et trois autres É.

5.10 TABLEAU DE RÉGIME [TR]

tr X = I = N
Y = II = à N, sur N

Ce champ décrit le régime syntaxique du mot-clé, c'est-à-dire la correspondance entre ses actants sémantiques, ses actants syntaxiques profonds et ses actants syntaxiques de surface. Par exemple, le régime ci-contre indique que le mot-clé (qui est biactanciel) exprime son premier actant sémantique (X) en tant que premier actant syntaxique profond (I), qui lui-même s'exprime, en surface, par un groupe nominal (N) ; le second actant sémantique du mot-clé (Y), s'exprime en tant que

second actant syntaxique profond (II), qui lui-même s'exprime, en surface, par un groupe prépositionnel en *à* ou en *sur*.

Noter que certaines lexies peuvent contrôler plus d'un régime, c'est-à-dire que leurs actants sémantiques peuvent s'exprimer par plus d'une configuration syntaxique profonde. Dans ce cas, chacun des régimes est introduit avec une numérotation l'identifiant de façon propre. Par exemple, APPLAUDIR**I** possède le régime suivant :

```
tr Rég. 1
   X = I = N
   Y = II = N
   Z = III = pour N [applaudir l'artiste pour sa performance]
   Rég. 2
   X = I = N
   Z = II = N [appplaudir la performance de l'artiste]
```

5.11 SYNONYMIE [SYN]

Le champ de synonymie contient les liens lexicaux apparentés à la synonymie. Il s'agit des quatre liens suivants, qui apparaissent entre accolades dans le DiCo — comme tout lien lexical standard ou non standard :

1. `{=}` : expression(s) constituée(s) du mot-clé accompagné d'un modificateur entièrement redondant ; par exemple, dans la fiche de BAGAGE**II** :

   `{=} bagage#II de connaissances, bagage#II intellectuel`

2. `{Syn}` ou `{QSyn}` : lexies liées au mot-clé par la fonction lexicale standard **`Syn`** (synonymie exacte ou approximative).

3. `{Cf}` : lexies sémantiquement très proches du mot-clé mais qui n'en sont pas des synonymes à proprement parler ; par exemple, dans la fiche de ABEILLE :

   `{Cf} frelon, guêpe; bourdon`

4. `{Gener}` : lexies liées au mot-clé par la fonction lexicale standard **`Gener`**.

5.12 FONCTIONS LEXICALES [FL]

Ce champ représente le cœur de la fiche DiCo. Il contient l'encodage, au moyen de fonctions lexicales (standard et non standard, paradigmatiques et syntagmatiques), des liens lexicaux contrôlés par le mot-clé — à l'exception des liens inclus dans le champ précédent. Voici, à titre d'exemple, le champ fonctions lexicales de la lexie À L'ARRIÈRE-PLAN :

```
fl {QAnti} _au premier plan_; _en avant_; _sur le devant de la scène_
     /*[X] se trouver À L'A.-P.*/
   {Oper1} être (relégué), figurer, se trouver [~]
     /*[X] commencer à se trouver À L'A.-P.*/
   {IncepOper1} passer, se retrouver [~]
     /*[Qqn./Qqch.] causer que X se trouve À L'A.-P.*/
   {CausOper1} faire passer, placer, mettre, reléguer [N=X ~]
     /*[Qqn./Qqch.] causer que X reste À L'A.-P.*/
   {CausContOper1} garder, maintenir [N=X ~]
```

Nous allons très brièvement expliquer certains formalismes utilisés pour l'encodage des liens de fonctions lexicales, en examinant les neuf points suivants :

1. formules de fonctions lexicales ;
2. gloses de « vulgarisation » des formules de fonctions lexicales ;
3. marques d'usage ;
4. séparateurs des différentes cibles des liens de fonctions lexicales ;
5. régimes des cibles ;
6. valeurs fusionnées ;
7. contraintes ;
8. exemples d'emploi ;
9. rubriques de fonctions lexicales.

5.12.1 Formules de fonctions lexicales

Les formules encodant les liens lexicaux au moyen des fonctions lexicales standard et non standard apparaissent dans les fiches de DiCo entre accolades — voir la sortie d'écran plus haut. Elles sont immédiatement suivies de la liste des cibles des liens en question.

Noter que la compilation du DiCo en DiCoSQL atomise les liens de fonctions lexicales. Par exemple, la première relation apparaissant plus haut se retrouve compilée sous forme des trois liens lexicaux distincts ci-contre.

QAnti	_au premier plan_
QAnti	_en avant_
QAnti	_sur le devant de la scène_

Dans le DiCo, les noms de locutions (phrasèmes complets dans la terminologie de la LEC) apparaissent entre deux tirets de soulignement.

5.12.2 Gloses de « vulgarisation » des formules de fonctions lexicales

Les noms des fonctions lexicales standard sont vulgarisés par des gloses apparaissant entre les marqueurs /*...*/ au-dessus du lien de fonction lexicale en question. Par exemple, le lien

de fonction lexicale `{Oper1}` a été vulgarisé plus haut par la formule `[X] se trouver À L'A.-P.`.

Lorsque la glose s'applique à plus d'un lien subséquent, le nombre de liens couverts est indiqué entre \…\ au tout début de la glose. Par exemple, on trouve dans l'article de APPLAUDISSEMENTS :

```
     /*\2\[Z] être la raison d'A.*/
{Oper3} provoquer, soulever, susciter [ART ~]
{Oper32} valoir [ART ~ à N=Y]
     /*\2\A. sont émis*/
{Func0} crépiter, résonner
{Magn.Func0} retentir
```

5.12.3 Marques d'usage

Les cibles des liens de fonctions lexicales peuvent être précédées d'une marque d'usage apparaissant entre guillemets : `"fam"`, `"vielli"`, `"angl"`, `"spéc"`, etc.

> La marque d'usage `"spéc"` signale que l'élément en question relève du langage d'un domaine de spécialité, donc d'une terminologie particulière. Pour l'instant, nous n'avons pas introduit de marque d'usage pour chacun des domaines spécifiques (domaine juridique, militaire, etc.).

5.12.4 Séparateurs des différentes cibles des liens de fonctions lexicales

Les différentes cibles d'un lien de fonction lexicale sont, par défaut, séparées par une virgule. Néanmoins, ces cibles peuvent manifester des écarts sémantiques notables. Elles sont alors séparées par un point-virgule. De plus, on peut observer, dans certains cas, une gradation d'intensité entre les cibles ; cette gradation est signalée par le séparateur « < ». Les données ci-dessous, extraites de la fiche de AVERSION, illustrent ces différents cas de figure :

```
     /*Intense*/
{Magn} grande | antépos, profonde, violente, vive < totale;
       insurmontable, irrépressible; immodérée | postpos,
       excessive
```

5.12.5 Régimes des cibles

Les régimes des cibles — c'est-à-dire les structures syntaxiques gouvernées par les cibles — sont indiqués entre crochets. Par exemple, pour la lexie BALAI :

```
{Real13} passer [le ~ Loc-in N=Z]
```

5.12.6 *Valeurs fusionnées*

Certaines valeurs de fonctions lexicales syntagmatiques sont fusionnées, c'est-à-dire qu'elles ne correspondent pas à des collocatifs du mot-clé, mais expriment à elles seules le sens du mot-clé et celui de la fonction lexicale syntagmatique. Les valeurs fusionnées sont précédées de l'opérateur de fusion « `//` ». Par exemple, pour la lexie ABOIEMENT, on trouve :

```
        /*Peu intense*/
    {AntiMagn} faible //jappement
```

On voit qu'une valeur fusionnée équivaut à un dérivé sémantique du mot-clé ; ici, JAPPEMENT est un **Syn** approximatif de ABOIEMENT. La fusion est donc un phénomène qui lie conceptuellement la notion de lien paradigmatique et celle de lien syntagmatique.

5.12.7 *Contraintes*

Certaines cibles de liens de fonctions lexicales possèdent des contraintes d'emploi. On doit notamment préciser pour les adjectifs s'ils sont antéposés ou postposés. Les contraintes sont introduites par l'opérateur « | ». Par exemple, on trouve dans ABOIEMENT :

```
        /*Indésirable*/
    {AntiBon} intempestifs | postpos et A. au pl
```

> Par défaut, les adjectifs français sont considérés dans le DiCo comme pouvant être à la fois antéposés et postposés. On ne spécifie donc les contraintes sur leur positionnement que lorsque seule une des deux positions est acceptable.

5.12.8 *Exemples d'emploi*

Il arrive que la mention des cibles de liens de fonctions lexicales soit accompagnée d'un exemple illustratif permettant de mieux comprendre le sens ou la forme exacte de la structure qui est encodée. Ces exemples apparaissent entre crochets **et** entre accolades (afin qu'ils se distinguent formellement des régimes). Par exemple, on trouve dans MEURTRE :

```
    {[X] essayer de faire croire
     qu'un M. est un N} maquiller [ART ~ en N] ["Il a maquillé ce
                        meurtre en accident/suicide."]
```

5.12.9 *Rubriques de fonctions lexicales*

On utilise parfois dans le champ de fonctions lexicales des fiches DiCo une ou plusieurs rubriques, qui regroupent des liens lexicaux reflétant une perspective particulière sur mot-vedette. Les noms de rubriques apparaissent entre accolades doubles. Par exemple, on trouve dans CIRE :

```
{{C. fabriquée par les abeilles}}

{=} cire d'abeille
{Petit réceptacle en C.
 que construisent les abeilles pour y stocker le miel [=f]} alvéole
     /*Bloc de f*/
{Multf} gaufre, rayon
```

5.13 EXEMPLES [EX]

Ce champ contient des exemples illustrant le comportement du mot-clé dans des phrases complètes. Les exemples sont normalement extraits de corpus linguistiques ou du Web ; souvent, on opère quelques ajustements pour que la phrase d'exemple fonctionne mieux de façon autonome, une fois déconnectée de son contexte textuel. Les exemples sont séparés par l'opérateur « | ». Voici, comme illustration, le champ exemple de la lexie CLOU :

```
ex  Des vêtements pendent à un clou planté dans le mur. | Une simple
    planche fixée par quelques clous obstruait l'entrée.
```

5.14 PHRASÉOLOGIE [PH]

Ce champ contient l'énumération des locutions dans lesquelles figure le mot-clé. Rappelons que les noms de locutions apparaissent dans le DiCo entre tirets de soulignement. Voici le champ phraséologie de CLOU.

```
ph  _au clou_
    _chaussettes à clous_
    _clou de girofle_
    _Des clous !_
    _enfoncer le clou_
    [maigre] _comme un clou <comme un cent de clous>_
    _ne pas en ficher un clou_
    _ne pas valoir un clou_
    _river son clou [à N]_
    _vieux clou_
```

5.15 AUTEUR [AUT]

Initiales du nom de la dernière personne qui a modifié la fiche.

5.16 DATE

Date de la dernière modification des informations contenues dans la fiche.

5.17 Heure

Heure à laquelle la fiche a été modifiée pour la dernière fois. La mise à jour de ce champ et du champ précédent est effectuée automatiquement par le logiciel FileMaker Pro®.

aut	APol	date	17/01/2005	12:06:15

6 Utilisation de l'interface DiCouèbe

6.1 Présentation générale de l'interface DiCouèbe

Le DiCouèbe est une interface, du type formulaire, permettant d'effectuer des requêtes prédéfinies sur une base de données SQL — voir Steinlin et autres (2004). La consultation du DiCouèbe se fait en ligne, au moyen d'un navigateur standard. Les résultats sont affichés dans des tables HTML. Il est possible d'effectuer les requêtes à l'aide de l'interface de requête ou directement en mode SQL.

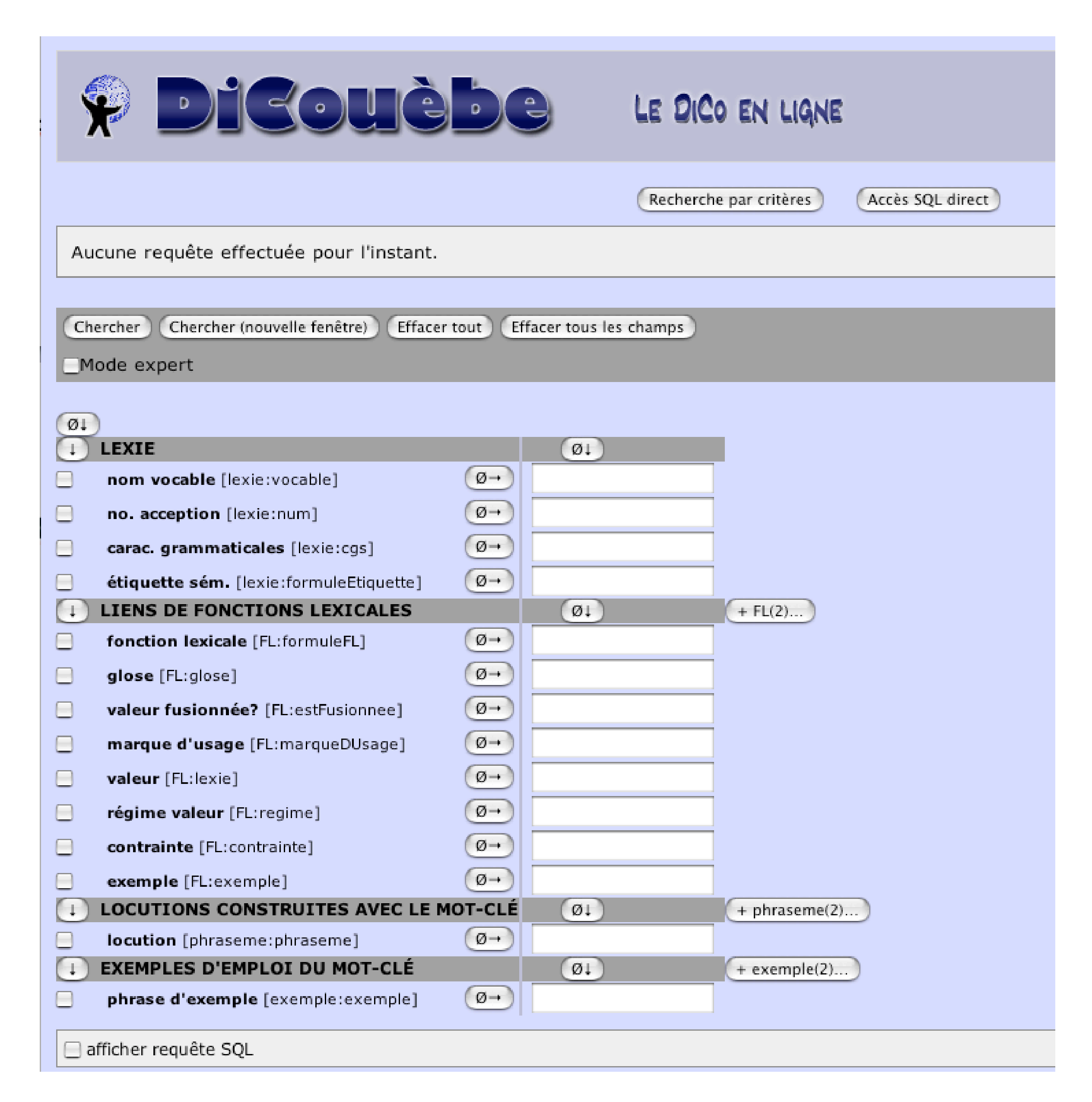
Il n'est pas possible pour l'instant d'exporter les tables de résultats sous un autre format à partir de l'interface. Cette fonctionnalité sera bientôt introduite.

Il existe deux modes de consultation du DiCoSQL au moyen du DiCouèbe :

1. Le **mode standard** est celui auquel on accède par défaut à partir de la page d'accueil du DiCouèbe[3]. Il donne accès aux champs centraux dans la structuration du réseau lexical du DiCoSQL.
2. Le **mode expert** donne accès à tous les champs du DiCoSQL, excepté le champ remarques (utile uniquement aux lexicographes du DiCo).

Le formulaire d'accès en mode standard est donné ci-dessous.

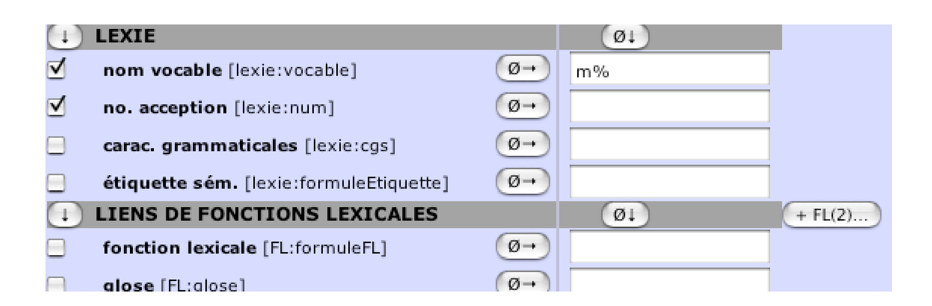
[3] http://www.olst.umontreal.ca/dicouebe



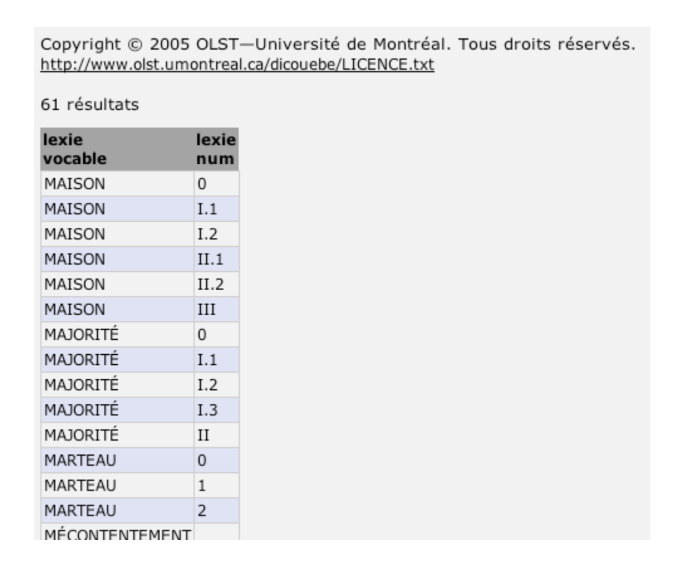
Comme on le voit, le formulaire d'interrogation standard est structuré en quatre sections, qui regroupent un nombre variable de types de données. Chaque section correspond à une **table** de données dans le DiCoSQL, et chaque type d'information accessible dans la section est un **champ** associé à cette table dans la base. À partir du mode expert, on peut accéder en tout à sept sections différentes (correspondant à sept tables du DiCoSQL).

Dans le formulaire, chaque type de donnée (chaque ligne interrogeable du formulaire) possède un nom en clair, par exemple **nom vocable**, suivi des noms entre crochets de la table et du champ correspondant dans la base DiCoSQL, par exemple **[lexie:vocable]** (c'est-à-dire, le champ `vocable` de la table `lexie`). Cette dernière information est utile à l'usager dans la mesure où, comme nous le verrons plus bas, c'est elle qui apparaît dans les en-têtes de colonnes dans les tableaux de résultats générés par le DiCouèbe.

Le mode d'utilisation du formulaire est très simple. On écrit dans les cases de droites les contraintes sur l'information devant apparaître dans le champ correspondant de la base et on clique à gauche dans les cases correspondant aux champs dont les valeurs doivent être affichées dans les tables de résultats. Par exemple, la sélection suivante permet d'afficher la liste de toutes les lexies (nom de vocable et numéro d'acception) dont le nom commence par la lettre « M », sachant que « % » est le caractère joker, utilisé dans les requêtes pour dénoter une série de caractères inconnus.



Cette requête donnera lieu à la génération de la table de résultats suivante :



Copyright © 2005 OLST—Université de Montréal. Tous droits réservés.
http://www.olst.umontreal.ca/dicouebe/LICENCE.txt

61 résultats

lexie vocable	lexie num
MAISON	0
MAISON	I.1
MAISON	I.2
MAISON	II.1
MAISON	II.2
MAISON	III
MAJORITÉ	0
MAJORITÉ	I.1
MAJORITÉ	I.2
MAJORITÉ	I.3
MAJORITÉ	II
MARTEAU	0
MARTEAU	1
MARTEAU	2
MÉCONTENTEMENT	

Nous allons maintenant décrire plus en détail le contenu du formulaire d'interrogation.

6.2 Structure du formulaire d'interrogation

Dans la description ci-dessous, nous faisons précéder du symbole « ‡ » les noms des données qui ne sont accessibles qu'en mode expert. En théorie, le nom en clair de chaque ligne du formulaire d'interrogation est suffisamment explicite pour que l'on puisse comprendre l'information à laquelle il fait référence, une fois que l'on connaît le contenu des différents champs d'une fiche DiCo (cf. section 5 plus haut).

Nous ne pouvons pas présenter ci-dessous les notions théoriques et descriptives sur lesquelles reposent les modélisations accessibles par le DiCouèbe. Certaines de ces notions ont été introduites dans les sections antérieures, d'autres seront présupposées connues par le lecteur. Cela est particulièrement vrai pour les données accessibles en mode expert — celles précédées du symbole « ‡ » dans les descriptions qui suivent.

6.2.1 Section « Lexie »

Cette section correspond à la table `lexie` du DiCoSQL et regroupe les données suivantes :

- **nom vocable**
- **no. d'acception**
- ‡**est acception de base ?** : valeur booléenne qui permet de déterminer si une lexie choisie est l'acception de base du vocable dont elle fait partie — `0` = non et `1` = oui.
- **carac. grammaticales**
- ‡**est étiquette ?** : valeur booléenne qui permet de déterminer si une lexie choisie est utilisée dans le DiCo comme étiquette sémantique — `0` = non et `1` = oui.
- ‡**nota bene**
- **étiquette sém.** : étiquette sémantique de la lexie dans le DiCo
- ‡**struct. actancielle**

Par exemple, si l'on demande l'affichage de toutes les données de la section « Lexie » pour les lexies du vocable SOUPÇON, on obtient la table de résultats suivante.

lexie vocable	lexie num	lexie ulb	lexie cgs	lexie estEtiquette	lexie nb	lexie formuleEtiquette	lexie formePropositionnelle
SOUPÇON	0	0	nom, masc	0			
SOUPÇON	I	1	nom, masc	0		fait de croire	: ~ DE LA personne X À PROPOS DE LA personne Y SUR LE fait Z (Y)
SOUPÇON	II	0	nom, masc, pas de pl, surtout indéf	0		petite quantité	: ~ DE X

6.2.2 Section « Liens de fonctions lexicales »

Cette section correspond à la table `FL` du DiCoSQL et regroupe les données suivantes :

- ‡**rubrique de fonction lexicale**
- **fonction lexicale**
- ‡**est standard ?**
- **glose**
- ‡**groupe valeur**
- ‡**gradation**

- **valeur fusionnée ?**
- **marque d'usage**
- **valeur**
- **régime valeur**
- **contrainte**
- **exemple**

Nous détaillons ci-dessous quelques-uns de ces types de données.

Rubrique de fonction lexicale

Nous avons déjà présenté dans la section 5 (voir 5.12.9) la notion de rubrique de fonctions lexicales. Dans le DiCoSQL, l'information correspondante est stockée de la façon suivante :

1. la valeur du nom de rubrique est `syn` pour les liens lexicaux du champ synonymie (voir plus haut section 5.11) ;
2. sa valeur est le nom de rubrique apparaissant dans la fiche DiCo lorsque le lien en question est encodé à l'intérieur d'une rubrique ;
3. il n'y a aucune valeur dans tous les autres cas.

Par exemple, si l'on fait afficher tous les noms de rubriques, de liens lexicaux et de cibles (avec leur régime) pour la lexie ABEILLE, on obtient le tableau de résultats suivant :

FL rubriqueFL	FL formuleFL	FL lexie	FL regime
syn	QSyn	cirage	
syn	QSyn	encaustique	
	PreparReal@	faire fondre	[ART ~]
	PreparReal@	pétrir	[ART ~]
	PreparReal@	travailler	[ART ~]
	enduire Labreal@1	enduire	[N=X avec/de ART ~]
	enduire Labreal@1	astiquer	[N=X à la ~]
	enduire Labreal@1	frotter	[N=X à la ~]
	enduire Labreal@1	astiquer	[N=X avec ART ~]
	enduire Labreal@1	frotter	[N=X avec ART ~]
	enduire Labreal@1	cirer	[N=X]
	A2 enduire Labreal@1	ciré	
	fabriquer Labreal@1	modeler	[N=X en ~]
	fabriquer en partie Labreal@1	à	[la ~]
	A2 fabriquer Labreal@1	de	[~]
	A2 fabriquer Labreal@1	en	[~]
	Qui a l'aspect <la consistance> de la C.	cireux	
C. fabriquée par les abeilles	=	cire d'abeille	
C. fabriquée par les abeilles	Petit réceptacle en C. que construisent les abeilles pour y stocker le miel	alvéole	
C. fabriquée par les abeilles	MultPetit réceptacle en C. que construisent les abeilles pour y stocker le miel	gaufre	
C. fabriquée par les abeilles	MultPetit réceptacle en C. que construisent les abeilles pour y stocker le miel	rayon	

Champs «est standard?» et «glose»

Le champ « est standard ? », indique si la fonction lexicale est standard ou non standard grâce à une valeur booléenne : « 0 » si la fonction lexicale n'est pas standard et « 1 » si elle l'est. Pour les fonctions standard, la glose de vulgarisation (voir section 5.12.2 ci-dessus) est donnée dans le champ « glose ».

lexie vocable	lexie num	FL formuleFL	FL estStandard	FL glose	FL lexie	FL regime
BALAI		QSyn	1		balai-brosse	
BALAI		QSyn	1		faubert	
BALAI		QSyn	1		vadrouille	
BALAI		Cf	1		brosse	
BALAI		Cf	1		plumeau	
BALAI		Cf	1		tête-de-loup	
BALAI		Cf	1		balayeuse	
BALAI		Cf	1		aspirateur	
BALAI		professionnel S1	1	Nom pour le professionnel X	balayeur	
BALAI		S2	1	Nom pour Y	poussière	
BALAI		S2	1	Nom pour Y	[des] saletés	
BALAI		S3	1	Nom pour Z	plancher	
BALAI		S3	1	Nom pour Z	sol	
BALAI		Magn	1	De grande taille	grand	
BALAI		Magn	1	De grande taille	long	
BALAI		AntiMagn	1	De petite taille	petit	
BALAI		AntiMagn	1	De petite taille	court	
BALAI		AntiMagn	1	De petite taille	balayette	
BALAI		Real13	1	[X] utiliser un B. pour nettoyer Z	passer	[le ~ Loc-in N=Z]
BALAI		Real12/3	1	[X] utiliser un B. pour nettoyer Z	balayer	[N=Y/Z]
BALAI		S0Real1	1	Nom pour [X] utiliser un B. pour nettoyer Z	_coup de ~_#I	
BALAI		S0Real1	1	Activité [X] utiliser un B. pour nettoyer Z	balayage	
BALAI		Partie d'un B.	0		brosse	[de ART ~]
BALAI		Partie d'un B.	0		manche	[à ~]
BALAI		Partie d'un B.	0		manche	[de ART ~]
BALAI		Endroit où l'on range les B.	0		placard	[à ~s]

Champs «groupe valeur» et «gradation»

Tel qu'expliqué plus haut dans la section 5.12.4, les différentes cibles d'un lien de fonction lexicale sont parfois séparées par des points-virgules, qui indiquent une distance sémantique importante, ou par l'opérateur « < », qui indique une gradation. On trouvera dans le champ « groupe valeur » un chiffre indiquant le regroupement sémantique auquel appartient chaque cible de lien de fonction lexicale. Le champ « gradation » précise si l'on distingue une ou plusieurs gradations entre les différentes cibles.

6.2.3 Section « Actants sémantiques »

Cette section correspond à la table `asemvar` du DiCoSQL, qui stocke l'information générée à partir de la formule propositionnelle du mot-clé (voir section 2.3). Elle n'est disponible qu'en mode expert et contient les informations suivantes :

- ‡**actant sémantique**
- ‡**dépendance** : indique l'actant sémantique éventuel de l'actant du mot-clé (voir exemple ci-dessous).
- ‡**optionnel** : précise si l'actant sémantique est optionnel ou non au moyen de booléens.
- ‡**étiquette sémantique**

Prenons comme illustration le cas de la lexie BAGUE1. Elle possède la formule propositionnelle suivante dans le DiCo :

```
: ~ DE L'individu X [QU'IL A À SON doigt Y(X)]
```

Celle-ci, une fois compilée, se retrouve encodée de la façon suivante dans la table `asemvar` du DiCoSQL :

asemvar asemvar	asemvar dependance	asemvar optionnel	asemvar formuleEtiquette
X		0	individu
Y	X	0	doigt

6.2.4 Section « Actants syntaxiques »

Cette section correspond à la table `asyntp` du DiCoSQL, qui stocke l'information générée à partir du schéma de régime du mot-clé (voir section 5.10). Comme la précédente, elle n'est disponible qu'en mode expert. Elle contient les informations suivantes :

- ‡**num. tableau** : indique de quel tableau de régime il s'agit, dans le cas où le mot-clé possède plusieurs régimes ; la valeur sera « `0` » en cas de régime unique.
- ‡**contraintes tableau** : en cas de régimes multiples, contient les éventuelles contraintes s'appliquant au régime en question.
- ‡**actant sémantique**
- ‡**actant syntaxique** : ce champ et le précédent permettent d'établir la correspondance entre actants sémantiques et actants syntaxiques profonds.
- ‡**marque d'usage** : indique la marque d'usage éventuelle qui s'applique à une réalisation de surface particulière.

- ‡**réalisation surface** : indique comment les actants syntaxiques profonds sont exprimés au niveau syntaxique de surface.
- ‡**contraintes surface**
- ‡**exemple**
- ‡**est obligatoire ?** : indique au moyen d'un booléen si l'actant en question est obligatoirement exprimé ou non.

Voici, à fin d'illustration, quelques régimes tels qu'encodés dans le DiCo et tels qu'ils apparaissent lorsque l'on interroge leur version compilée dans le DiCoSQL à partir du DiCouèbe.

<u>Lexie DÉFAITE</u>

Données du DiCo :

```
tr  X = I = de N, A-poss, Adj [défaite française]
    Y = II = _face à_ N, "litt" _aux mains de_ N
    Z = III = dans N
```

Données extraites du DiCoSQL :

lexie vocable	lexie num	asyntp numTableau	asyntp contrainteTableau	asyntp asem	asyntp asynt	asyntp marqueDUsage	asyntp syntagme	asyntp contraintes	asyntp exemple	asyntp estObligatoire
DÉFAITE		0		X	I		de N			0
DÉFAITE		0		X	I		A-poss			0
DÉFAITE		0		X	I		Adj		[défaite française]	0
DÉFAITE		0		Y	II		_face à_ N			0
DÉFAITE		0		Y	II	"litt"	_aux mains de_ N			0
DÉFAITE		0		Z	III		dans N			0

<u>Lexie COUP DE FOUDRE1</u>

Données du DiCo :

```
tr  Rég. 1
    X = I = de N, A-poss
    Y = II = pour N
    Rég. 2 | C. D. F. est réciproque
    X+Y = I = entre N et N, A-poss
```

Données extraites du DiCoSQL :

asyntp numTableau	asyntp contrainteTableau	asyntp asem	asyntp asynt	asyntp marqueDUsage	asyntp syntagme	asyntp contraintes	asyntp exemple	asyntp estObligatoire
1		X	I		de N			0
1		X	I		A-poss			0
1		Y	II		pour N			0
2	C.D.F.estréciproque	X + Y	I		entre N et N			0
2	C.D.F.estréciproque	X + Y	I		A-poss			0

6.2.5 Section « Locutions construites avec le mot-clé »

Cette section correspond à la table phraseme du DiCoSQL, qui stocke l'information générée à partir du schéma de champ phraséologie de la fiche DiCo (voir section 5.14). Elle contient les informations suivantes :

- **‡marque d'usage**
- **locution**
- **‡base (si collocatif)** : précise le nom des bases qui contrôlent le choix de la locution lorsque celle-ci ne s'emploie que comme collocatif au sein de collocations.
- **‡exemple**

Voici quelques exemples des données contenues dans le DiCoSQL, une fois interrogées au moyen du DiCouèbe.

Lexie FLÈCHE**I**

Données du DiCo :

```
ph  _comme une flèche_
    [“Lorsque le cours se termine tard à midi, il court à l'autobus comme
    une flèche, l'estomac vide et le dos un peu courbé.”]
    _en flèche_
    [“Père a décidé de doubler sa production agricole, car les prix
    montent en flèche.”]
    _faire flèche de tout bois_
    _flèches de l'amour_
```

Données extraites du DiCoSQL :

phraseme marqueDUsage	phraseme phraseme	phraseme base	phraseme exemple
	comme une flèche		[“Lorsque le cours se termine tard à midi, il court à l'autobus comme une flèche, l'estomac vide et le dos un peu courbé.”]
	en flèche		[“Père a décidé de doubler sa production agricole, car les prix montent en flèche.”]
	faire flèche de tout bois		
	flèches de l'amour		

Lexie GUICHET**1**

Données du DiCo :

ph	[jouer] _à guichets fermés_

Données extraites du DiCoSQL :

phraseme marqueDUsage	phraseme phraseme	phraseme base	phraseme exemple
	à guichets fermés	[jouer]	

6.2.6 Section « Exemples d'emploi du mot-clé »

Cette section correspond à la table `exemple` du DiCoSQL, qui stocke l'information générée à partir du champ exemples de la fiche DiCo (voir section 5.13). Elle ne contient qu'un seul type d'information :

- **exemple**

lexie vocable	lexie num	exemple exemple
COUP DE FOUDRE	1	Entre Fay et Tom, ça a été le coup de foudre immédiat.
COUP DE FOUDRE	2	En arrivant au Canada, il a eu le coup de foudre pour ce pays.
COUP DE FOUDRE	2	Les visiteurs de la galerie sont surtout à la recherche d'un coup de foudre, du tableau qui va les faire craquer.

6.2.7 Section « Info article »

Cette section ne s'affiche qu'en mode de consultation expert. Elle correspond à la table `article` du DiCoSQL, qui stocke de l'information « éditoriale » sur la fiche DiCo d'où sont extraites les données :

- **‡statut**
- **‡date modif.** : non accessible pour l'instant.
- **‡heure modif.** : non accessible pour l'instant.
- **‡auteur**

Nous allons conclure cette section sur l'interface DiCouèbe en montrant comment on peut pratiquement utiliser les DiCouèbe pour extraire de l'information de la base DiCoSQL.

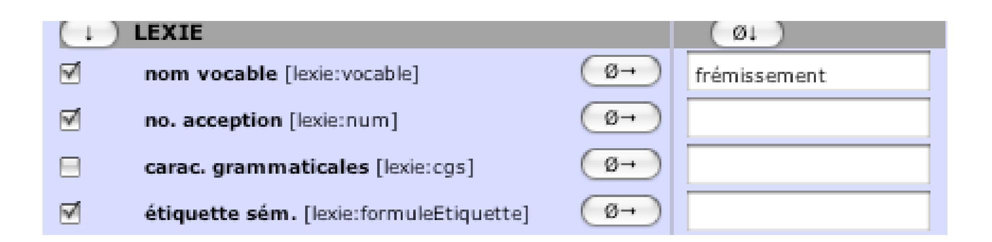
6.3 UTILISATION DU DICOUÈBE

6.3.1 Principes généraux d'interrogation

Tel que mentionné plus haut, une série de cases à cocher, dans la colonne de gauche du formulaire d'interrogation, permet de choisir les éléments que l'on veut obtenir lors de

l'affichage du résultat. Les cases de la colonne de droite permettent de contraindre les requêtes.

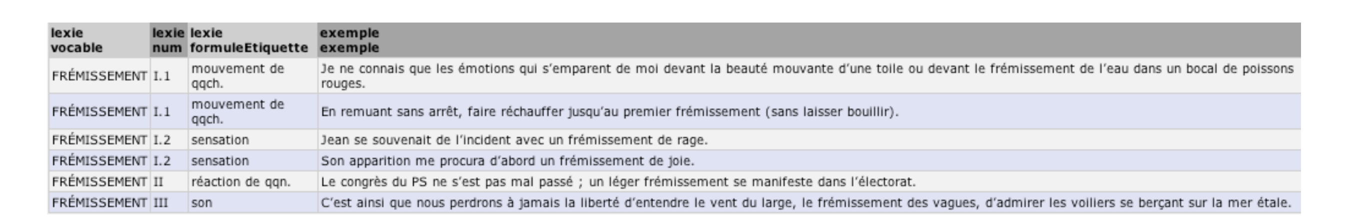
Ainsi, pour obtenir la liste de toutes les lexies du vocable FRÉMISSEMENT en faisant afficher leur nom au complet (nom de vocable et numéro d'acception) avec leur étiquette sémantique, il faut compléter le formulaire de la façon suivante :



Une fois la requête évaluée, on obtient le tableau de résultats ci-dessous :

lexie vocable	lexie num	lexie formuleEtiquette
FRÉMISSEMENT	0	
FRÉMISSEMENT	I.1	mouvement de qqch.
FRÉMISSEMENT	I.2	sensation
FRÉMISSEMENT	II	réaction de qqn.
FRÉMISSEMENT	III	son

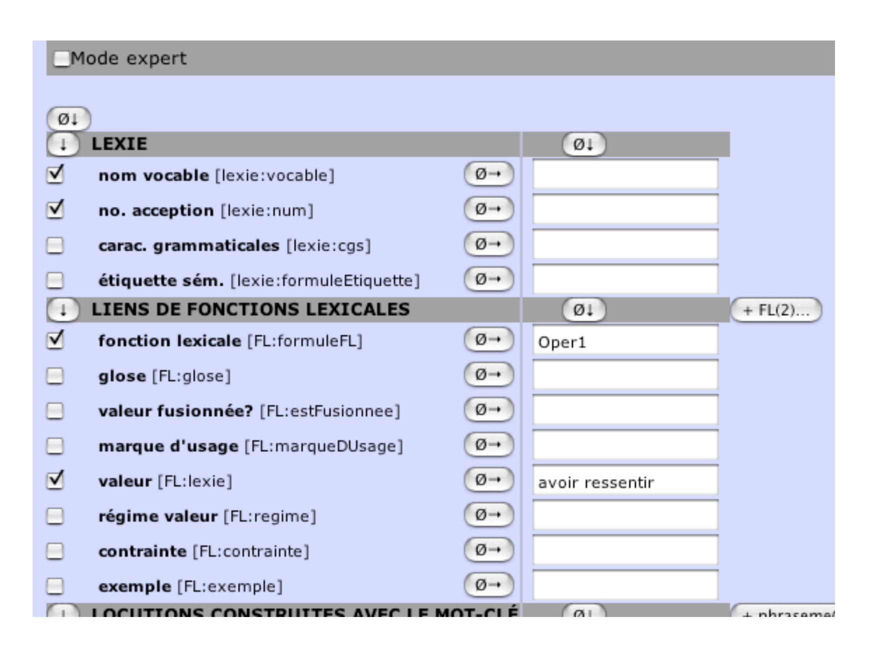
Si l'on trouve que le tableau n'est pas assez explicite, on peut le compléter en utilisant la même requête que précédemment, mais en cochant en plus la ligne *phrase d'exemple* de la section « Exemples d'emploi du mot-clé » du formulaire. Cela donne comme résultat :



lexie vocable	lexie num	lexie formuleEtiquette	exemple exemple
FRÉMISSEMENT	I.1	mouvement de qqch.	Je ne connais que les émotions qui s'emparent de moi devant la beauté mouvante d'une toile ou devant le frémissement de l'eau dans un bocal de poissons rouges.
FRÉMISSEMENT	I.1	mouvement de qqch.	En remuant sans arrêt, faire réchauffer jusqu'au premier frémissement (sans laisser bouillir).
FRÉMISSEMENT	I.2	sensation	Jean se souvenait de l'incident avec un frémissement de rage.
FRÉMISSEMENT	I.2	sensation	Son apparition me procura d'abord un frémissement de joie.
FRÉMISSEMENT	II	réaction de qqn.	Le congrès du PS ne s'est pas mal passé ; un léger frémissement se manifeste dans l'électorat.
FRÉMISSEMENT	III	son	C'est ainsi que nous perdrons à jamais la liberté d'entendre le vent du large, le frémissement des vagues, d'admirer les voiliers se berçant sur la mer étale.

6.3.2 Requêtes avec disjonctions

On peut introduire des disjonctions dans les requêtes en inscrivant dans les cases du formulaire plusieurs éléments séparés par des blancs. Ainsi, la requête suivante permet d'obtenir la liste de toutes les lexies qui ont soit *avoir* soit *ressentir* comme cible du lien de fonction lexicale $\mathbf{Oper}_1$:



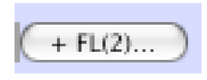
Voici le début de la table de résultats correspondante :

86 résultats

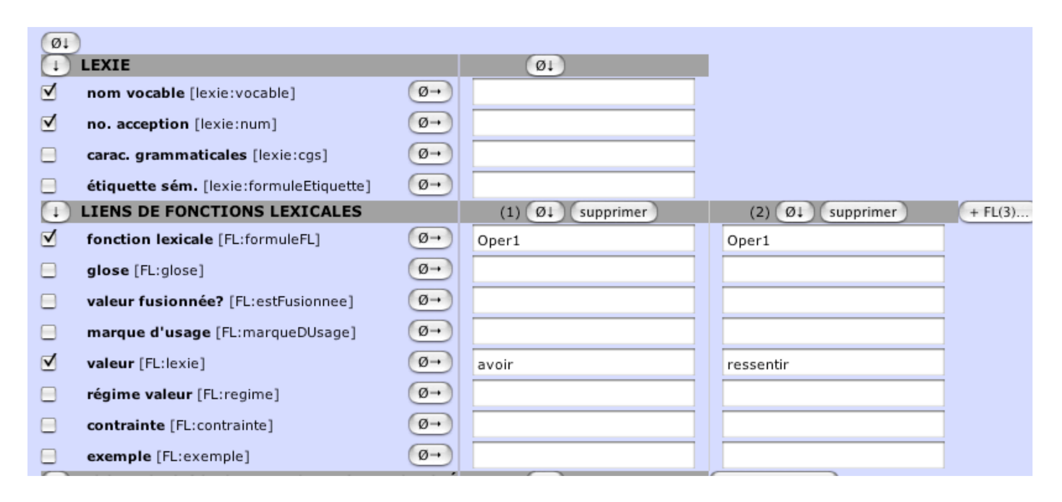
lexie vocable	lexie num	FL formuleFL	FL lexie
ANGOISSE		Oper1	avoir
ANGOISSE		Oper1	ressentir
BAGAGE	I.1	Oper1	avoir
BAGAGE	II	Oper1	avoir
BAGUE	2	Oper1	avoir
BARBE		Oper1	avoir
BON SENS		Oper1	avoir
CHIEN	I	Oper1	avoir
CONVERSATION	a	Oper1	avoir
COUVERCLE		Oper1	avoir
CRAINTE		Oper1	avoir
CRAINTE		Oper1	ressentir
DÉCEPTION		Oper1	ressentir
DÉSIR	1	Oper1	avoir
DÉSIR	1	Oper1	ressentir
EFFROI		Oper1	ressentir
ÉLOQUENCE		Oper1	avoir
ÉMOI		Oper1	ressentir
ENTREVUE		Oper1	avoir
ÉTAGE	1	Oper1	avoir
ÉTAGE	2	Oper1	avoir
FACTEUR(3)	II.1	Oper1	avoir

6.3.3 Requêtes avec conjonctions

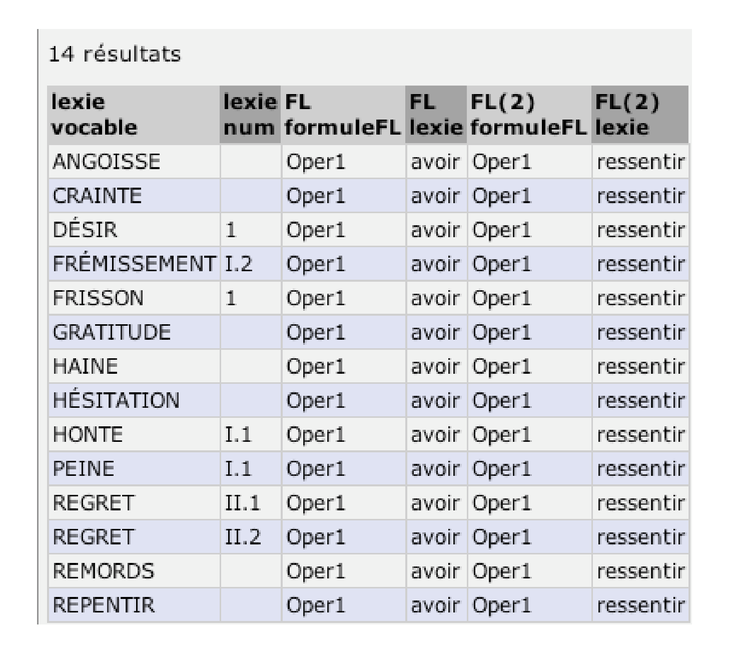
Toutes les sections, excepté la première (**Lexie**), peuvent être dédoublées afin d'effectuer des requêtes avec conjonctions. Cela se fait en cliquant sur les boutons du type ci-contre.



Par exemple, on lance la requête suivante pour obtenir la lite de toutes les lexies qui ont *avoir* **et** *ressentir* comme cibles du lien de fonction lexicale **`Oper1`** :



Cette requête donne évidemment beaucoup moins de résultats que la requête disjonctive décrite dans la section précédente. On obtient en effet le tableau de résultats ci-dessous :

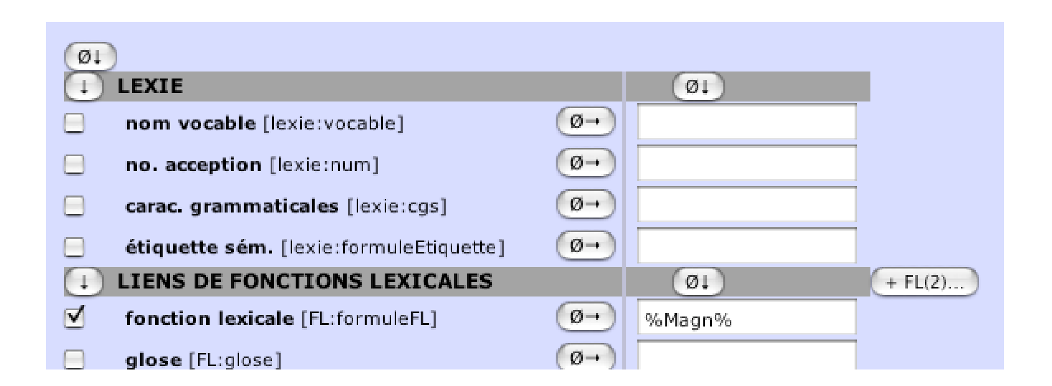


14 résultats

lexie vocable	lexie num	FL formuleFL	FL lexie	FL(2) formuleFL	FL(2) lexie
ANGOISSE		Oper1	avoir	Oper1	ressentir
CRAINTE		Oper1	avoir	Oper1	ressentir
DÉSIR	1	Oper1	avoir	Oper1	ressentir
FRÉMISSEMENT	I.2	Oper1	avoir	Oper1	ressentir
FRISSON	1	Oper1	avoir	Oper1	ressentir
GRATITUDE		Oper1	avoir	Oper1	ressentir
HAINE		Oper1	avoir	Oper1	ressentir
HÉSITATION		Oper1	avoir	Oper1	ressentir
HONTE	I.1	Oper1	avoir	Oper1	ressentir
PEINE	I.1	Oper1	avoir	Oper1	ressentir
REGRET	II.1	Oper1	avoir	Oper1	ressentir
REGRET	II.2	Oper1	avoir	Oper1	ressentir
REMORDS		Oper1	avoir	Oper1	ressentir
REPENTIR		Oper1	avoir	Oper1	ressentir

6.3.4 Caractère joker

Nous avons déjà mentionné qu'un caractère joker — « `%` » — peut être utilisé dans les requêtes. Ainsi, la requête suivante fait afficher la liste de tous les noms de fonctions lexicales contenant la chaîne de caractères « `Magn` » :



Ceci termine notre brève présentation du DiCo et du DiCouèbe. Nous sommes bien entendu conscients du caractère partiel de cette présentation, mais nous espérons qu'elle sera suffisante pour aider le lecteur à se familiariser avec la ressource qui lui est offerte. Nous recevrons avec plaisir toutes suggestions ou critiques qui pourraient nous guider dans la préparation des versions futures de ce document. Les auteurs peuvent être contactés aux adresses suivantes : anne.laure.jousse@umontreal.ca et alain.polguere@umontreal.ca.

7 Glossaire de quelques notions de base

Base (d'une collocation). Élément qui retient son sens dans la collocation et la contrôle fonctionnellement.

Collocatif (de la base). Élément d'une collocation fonctionnellement contrôlé par la base, c'est-à-dire dont le choix est contraint par celle-ci.

Collocation. Expression semi-idiomatique.

Combinatoire lexicale restreinte. Ensemble des propriétés d'une lexie qui contrôle sa capacité de se combiner avec d'autres lexies et n'est déterminé ni par son sémantisme ni par ses propriétés grammaticales.

DEC. *Dictionnaire explicatif et combinatoire* : dictionnaire « théorique » postulé par la lexicologie explicative et combinatoire.

Dérivation sémantique. Relation sémantique forte entre lexies qui tend à se manifester morphologiquement en langue. La dérivation morphologique est un cas particulier de dérivation sémantique.

Fonction lexicale. Entité lexicale qui est une généralisation de la notion de lexie et dénote un « potentiel de lexies ». Les fonctions lexicales sont utilisées dans le DiCo pour modéliser les liens lexicaux paradigmatiques et syntagmatiques.

LEC. *Lexicologie explicative et combinatoire* : composante lexicologique de la théorie Sens-Texte.

Lexie. Unité lexicale.

Locution. Expression idiomatique qui est une lexie à part entière. Les locutions sont aussi appelées *phrasèmes complets* dans la lexicologie explicative et combinatoire.

LAF. *Lexique actif du français* : dictionnaire grand public construit à partir des données du DiCo.

Mot-clé [d'une fiche de DiCo]. Lexie dont la fiche DiCo fait la description.

8 Références bibliographiques

Mel'čuk I. (1996) Lexical Functions : A Tool for the Description of Lexical Relations in a Lexicon. In L. Wanner (réd.) : *Lexical Functions in Lexicography and Natural Language Processing*, Amsterdam/Philadelphia, Benjamins, 37-102.

Mel'čuk I. et autres (1984, 1988, 1992, 1999) *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques, vol. I-IV*, Montréal, Les Presses de l'Université de Montréal.

Mel'čuk I., Clas A., Polguère A. (1995) *Introduction à la lexicologie explicative et combinatoire*, Louvain-la-Neuve, Duculot.

Mel'čuk I., Polguère A. (en préparation) *Lexique actif du français*.

Polguère A. (sous presse) Lexical function standardness. In L. Wanner (réd.) : *Festschrift in Honour of Igor Mel'čuk*, Amsterdam/Philadelphia, John Benjamin.

Polguère A. (à paraître) *Soleil insoutenable* et *chaleur de plomb* : le statut linguistique des greffes collocationnelles.

Polguère A. (2003a) *Lexicologie et sémantique lexicale. Notions fondamentales*, coll. « Paramètres », Montréal, Les Presses de l'Université de Montréal.

Polguère A. (2003b) Étiquetage sémantique des lexies dans la base de données DiCo. *Traitement Automatique des Langues* (*T.A.L.*), 44:2, 39-68.

Polguère A. (2000) Towards a theoretically-motivated general public dictionary of semantic derivations and collocations for French. *Proceedings of EURALEX'2000*, Stuttgart, 517-527.

Steinlin J., Kahane S., Polguère A., El Ghali A. (2004) De l'article lexicographique à la modélisation objet du dictionnaire et des liens lexicaux. *Actes de EURALEX'2004*, Lorient (France), 177-186.

Annexe : exemples de liens de fonctions lexicales standard

Chaque fonction lexicale **f** est illustrée par une application donnée

f(*<mot-clé>*) = *<valeur>*

La liste ci-dessous est ordonnée selon l'ordre canonique d'encodage des liens lexicaux standard dans les fiches DiCo.

QSyn(*naïf*)	= *candide*
Gener(*marteau*)	= *outil*
Conv$_{321}$(*vendre*)	= *acheter*
Anti(*bien*)	= *mal*
Contr(*mer*)	= *terre*
Epit(*gagnant*)	= *heureux* | prépos
Figur(*jalousie*)	= *démon* [*de la* ~]
Sing(*neige*)	= *flocon* [*de* ~]
Mult(*navire*)	= *flotte*
Cap(*association*)	= *président* [*de* ART ~]
Equip(*navire*)	= *équipage* [*de* ART ~]
Germ(*colère*)	= *ferment* [de la ~]
Centr(*nuit*)	= *coeur* [*de la* ~]
Culm(*colère*)	= *paroxysme* [*de* ART ~]
V$_0$(*serment*)	= *jurer*
S$_0$(*chanter*)	= *chant*
A$_0$(*fleuve*)	= *fluvial*
Adv$_0$(*goulu*)	= *goulûment*
Able$_1$(*nuire*)	= *nocif*
Qual$_2$(*tromper*)	= *naïf*
Magn(*pleurer*)	= *comme une Madeleine*
IncepPredPlus(*fièvre*)	= *augmenter*

Ver(*explication*)	=	*claire*
Bon(*se porter*)	=	*comme un charme*
IncepPredPejor(*mœurs*)	=	*se relâcher*
$\mathbf{Pos}_2$(*préjugé*)	=	*favorable* | postpos
$\mathbf{Adv}_1$(*mépris*)	=	*avec* [(ART) ~]
Instr(*téléphone*)	=	*à* [*le* ~], *par* [~]
$\mathbf{Loc}_{in}$(*campagne*)	=	*à* [*la* ~]
Propt(*maladie*)	=	*pour cause* [*de* ~]
$\mathbf{Oper}_1$(*suggestion*)	=	*faire* [ART~]
$\mathbf{Func}_0$(*entrevue*)	=	*se dérouler*
$\mathbf{Labor}_{12}$(*interrogatoire*)	=	*soumettre* [N_Y *à* ART ~]
$\mathbf{Real}_1$(*parachute*)	=	*sauter* [*en* ~]
$\mathbf{Fact}_0$(*désir*)	=	*se réaliser*
$\mathbf{Labreal}_{12}$(*guichet* [*automatique*])	=	*retirer* [N_Y *à* ART ~]
Involv(*plaque*)	=	*recouvrir* [N]
Manif(*sourire*)	=	*se dessiner*, *s'esquisser* [*sur* ART *visage*]
Prepar(*drap*)	=	*mettre* [ART ~]
$\mathbf{ProxOper}_1$(*désespoir*)	=	*être au bord* [*du* ~]
Degrad(*lait*)	=	*tourner*
Son(*éléphant*)	=	*barrir*
Imper(*se coucher*)	=	« *Au lit !* »
$\mathbf{ResultOper}_1$(*victoire*)	=	*avoir* [*la* ~]
Obstr(*vue*)	=	*se brouiller*
Excess(*yeux*)	=	*sortir de leurs orbites*
Stop(*moteur*)	=	*caler*
$\mathbf{Excess}_{color}$(*visage*)**.**$\mathbf{Sympt}_{13}$(*colère*)	=	*être rouge de colère*